

RESEARCH

Open Access



# DFA-Net: Dual multi-scale feature aggregation network for vessel segmentation in X-ray digital subtraction angiography

He Deng<sup>1</sup>, Xu Liu<sup>1</sup>, Tong Fang<sup>1</sup>, Yuqing Li<sup>1\*</sup> and Xiangde Min<sup>2\*</sup>

\*Correspondence:

liyuqing@wust.edu.cn;  
minxiangde0129@126.com

<sup>1</sup> School of Computer Science and Technology, Hubei Province Key Laboratory of Systems Science in Metallurgical Process, Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System, Wuhan University of Science and Technology, Wuhan 430081, China

<sup>2</sup> Department of Radiology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China

## Abstract

Even though deep learning is fascinated in fields of coronary vessel segmentation in X-ray angiography and achieves prominent progresses, most of those models probably bring high false and missed detections due to indistinct contrast between coronary vessels and background, especially for tiny sub-branches. Image improvement technique is able to better such contrast, while boosting extraneous information, e.g., other tissues with similar intensities and noise. If incorporating features derived from original and enhanced images, the segmentation performance is improved because those images comprise complementary information from different contrasts. Accordingly, inspired from advantages of contrast improvement and encoding-decoding architecture, a dual multi-scale feature aggregation network (named DFA-Net) is introduced for coronary vessel segmentation in digital subtraction angiography (DSA). DFA-Net integrates the contrast improvement using exponent transformation into a semantic segmentation network that individually accepts original and enhanced images as inputs. Through parameter sharing, multi-scale complementary features are aggregated from different contrasts, which strengthens leaning capabilities of networks, and thus achieves an efficient segmentation. Meanwhile, a risk cross-entropy loss is enforced on the segmentation, for availably decreasing false negatives, which is incorporated with Dice loss for joint optimization of the proposed strategy during training. Experimental results demonstrate that DFA-Net can not only work more robustly and effectively for DSA images under diverse conditions, but also achieve better performance, in comparison with state-of-the-art methods. Consequently, DFA-Net has high fidelity and structure similarity to the reference, providing a way for early diagnosis of cardiovascular diseases.

**Keywords:** Deep learning, Coronary vessel segmentation, Digital subtraction angiography

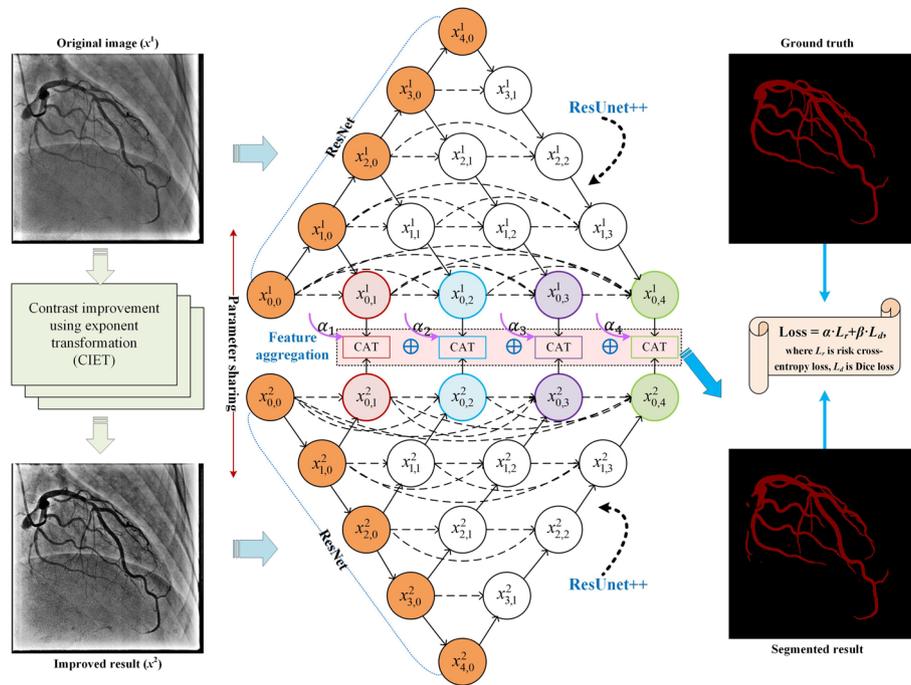
## Introduction

Cardiovascular diseases (CVDs), the top global cause of death, represented 32% of all worldwide deaths (about 17.9 million) in 2019, in which 85% were due to heart attack and stroke [1]. Of these diseases, coronary artery diseases (CADs) are the most common [2], which potentially leads to the sudden death of patients [3]. Because of plaque

build-up in coronary arteries, the narrowing (or stenosis) of the lumen is a proeguminal cause of CADs, which restricts blood flow to cardiac muscle, depriving the heart of oxygen and nutrient supplements, ultimately inducing myocardial ischemia and infarction [4]. At present, X-ray digital subtraction angiography (DSA) is considered as the gold standard in clinical diagnosis and interventional treatment of CADs [2], because such tactic visualizes the severity of vessel narrowing, protrusions, bifurcation, stenosis, etc.. Under this case, accurate and efficient segmentation of coronary vessels is a crucial issue, which provides the basis for the quantification and assessment of vascular stenosis [5, 6]. Nevertheless, manual delineation is expensive, and with high inter- and intra-reader variations [7]. On the other hand, fully automatic tactics eliminate potential subjectivity and supply quantitative and systematic measure of diameter reduction [5, 6], but such strategy is challenging owing to low resolution, high levels of noise, and complex structures with different vessel shapes and tiny sub-branches [8].

Automatic segmentation methods of coronary vessels are roughly divided into traditional and learning-based techniques [9]. Conventional algorithms are usually based on pattern recognition, model, and tracking. Even although those strategies produce promising results, they are not robust under complicate conditions, e.g., low contrast / resolution, heavy noise, and nonuniform intensity distributions of background [8]. Recently, with advances in deep learning, learning-based techniques achieve superior segmentation performance for coronary angiographic images in terms of accuracy and time-consuming [10–12], and such tactic generally generates a prediction map of pixel-wise probability. Among them, U-shaped architecture and skip-connections (e.g., U-net) exhibit great advantages in medical image segmentation due to strong capability of feature extraction [13]. Accordingly, lots of segmentation networks have been presented according to the basic U-net structure, e.g., U-net++ [14], attention U-net [15], and U-net 3+ [16]. Although U-net and its variants efficiently balance context and local accuracy, they are difficult to explicitly model long-range contextual interactions owing to the intrinsic locality of convolution (conv), which probably misses global contextual information in the extracted features [12]. To address the shortcomings, Dong et al. proposed a new multi-attention, multi-scale 3D convolutional neural network (CNN) for coronary artery segmentation [17]. Shi et al. introduced an affinity feature strengthening network by jointly modelling the geometry and refining pixel-wise segmentation features using a contrast-insensitive, multiscale affinity strategy [18]. Despite many efforts are dedicated to perfect segmentation performance of vessels, the performance is still unsatisfactory, and further improvements are required in such task.

It is known that DSA is a fluoroscopic technique that generates images of blood vessels without interfering shadows from overlapping tissues. Through the administered intensifier (e.g., contrast medium) in X-ray coronary angiogram, the structures with the contrast medium passing appear dark gray, while the background appears light gray (see Fig. 1). If the contrast between foreground objects (e.g., coronary vessels) and background (e.g., other tissues) is higher, the better segmentation result will be achieved. Nevertheless, due to low resolution and heavy noise, such contrast is usually inconspicuous, especially on tiny sub-branches, which probably leads to inefficient segmentation performance, e.g., high false and/or missed detections. This implies that the contrast improvement has a potential to perfect the segmentation



**Fig. 1** Framework of the proposed DFA-Net for coronary vessel segmentation in X-ray angiography, in which an original DSA image and its improved result by CIET are fed into ResUNet++ respectively, followed by the aggregation of dual multi-scale features, and the segmentation is acquired accordingly. The model is optimized by minimizing the joint risk cross-entropy loss and Dice loss synchronously

performance of coronary vessels. However, such way also enhance additionally extraneous information, e.g., other tissues with similar intensities and noise (see Fig. 1). Under this case, an incorporation of original and enhanced images can improve the performance of vessel segmentation networks, because those images embed complementary information from different contrasts. Inspired from merits of contrast enhancement (e.g., Retinex decomposition [19]) and encoding-decoding structure (U-net and its variants), a dual multi-scale feature aggregation network (termed as DFA-Net) is proposed for coronary vessel segmentation in DSA images. DFA-Net incorporates the contrast improvement using exponent transformation (CIET) into a semantic segmentation network (e.g., ResUNet++ [20]) that individually accepts original and contrast-improved images as inputs. Via parameter sharing and dual multi-scale feature aggregation of complementary information derived from different contrasts, an effective segmentation is achieved accordingly. Moreover, DFA-Net is optimized by jointly minimizing the risk cross-entropy loss and Dice loss. The contributions of this work are summarized as follows.

1. Inspired from merits of contrast enhancement and encoding-decoding architecture, a dual multi-scale feature aggregation network (DFA-Net) is proposed, which individually takes original and enhanced images as inputs for decoding complementary information from different contrasts, and outputs effective predictions of coronary vessels through the minimum of joint risk cross-entropy loss and Dice loss.

2. To address challenges specific to the contrast enhancement between coronary vessels and complex background, CIET is proposed, which adopts an exponent to substitute for logarithm in traditional multi-scale Retinex decomposition for avoiding ambiguity and negative values. Incorporating original images, multi-scale features leaning capability of network is reinforced, which guarantees an active segmentation.
3. Considering different risks between false negatives and false positives, a risk cross-entropy loss is used to availably lessen false negatives, which is integrated with Dice loss for jointly optimizing DFA-Net during training. Experimental results prove that DFA-Net can not only work more robustly and effectively, but also achieve better performance regarding five quantitative measures, compared with state-of-the-art methods.

## Proposed method

### Framework of DFA-Net

The framework of DFA-Net is showed in Fig. 1, a dual branch structure, which respectively accepts original and enhanced images as input, inspired from advantages of contrast improvement (CIET) and encoding-decoding architecture (ResUnet++). To start with, the contrast between foreground objects and background in an original DSA image is enhanced through CIET. Since CIET also enlarges irrelevant content, e.g., other tissues with similar intensities and noise, original and enhanced images are individually input to ResUnet++, for decoding multi-scale complementary information from different contrasts. After the aggregation (concatenation and addition) of dual multi-scale features under adaptive weighting parameters, the segmentation is acquired accordingly, in which parameter sharing is used during training. Meanwhile, DFA-Net is optimized by minimizing the joint risk cross-entropy loss and Dice loss synchronously.

### Contrast improvement

Low contrast between coronary vessels and background undoubtedly brings some difficulties in DSA segmentation. Derived from the multi-scale Retinex decomposition [19], CIET is proposed to enhance the contrast. For a given DSA image  $\mathbf{I}$ , Retinex decomposition is described as,

$$\mathbf{R} = \sum_{i=1}^N \omega_i \log \left( \frac{\mathbf{I}}{\mathbf{G}_i * \mathbf{I}} \right), \text{ where } \sum_{i=1}^N \omega_i = 1, \quad (1)$$

where  $\mathbf{R}$  is enhancement result,  $\mathbf{G}_i$  is the  $i^{\text{th}}$  Gaussian filtering function with standard derivation  $\sigma_i$  (named scale parameter),  $N$  is the number of scale parameters,  $*$  is the conv operator, and  $\omega_i$  is the weighting of  $i^{\text{th}}$  scale parameter, respectively.

In Eq. (1), the logarithm potentially yields negative values, which causes ambiguity in the transformation. Thus, an exponent function is applied to replace that logarithm, which is defined as,

$$S(x) = e^{1 - \frac{2}{e^{\alpha x} + 1}} - 1, x \in [0, 1], \text{ where } \alpha = \ln \frac{1 + \ln 2}{1 - \ln 2}, \quad (2)$$

The weighting parameter  $\alpha$  guarantees that the codomain of  $S(x)$  is in the interval  $[0, 1]$ , where  $S(0) = 0$  and  $S(1) = 1$ . The function  $S(x)$  is monotonically increasing in the range of  $[0, 1]$ , avoiding the ambiguity and negative values in the transformation. The graph of  $y = x - S(x), x \in [0, 1]$  is showed in Fig. 2, where 0,  $x_0$  and 1 are three zeros. It can be found that  $S(x) < x$ , if  $x \in (0, x_0)$ , and  $S(x) > x$ , if  $x \in (x_0, 1)$ . Under this case, when the intensity of an image is normalized to  $[0, 1]$ , in those pixels with low magnitude ( $< x_0$ ), the intensity is expanded. In contrast, high intensity ( $> x_0$ ) is suppressed. This property is used to stretch the intensity distribution of an image, which is favorable to better the image contrast.

Considering Eq. (2), Eq. (1) becomes as,

$$\mathbf{R}' = \sum_{i=1}^N \omega_i \mathbf{S} \left( \text{norm} \left( \frac{\mathbf{I}}{\mathbf{G}_i * \mathbf{I}} \right) \right), \text{ where } \sum_{i=1}^N \omega_i = 1, \tag{3}$$

where  $\mathbf{S}$  denotes the exponent function, and  $\text{norm}$  denotes a normalization operation.

To balance the enhancement performance against computation complexity of Gaussian filtering, the parameters  $\sigma_i$  in Eq. (3) are chosen as the large-, middle- and small-scale parameters, that is,

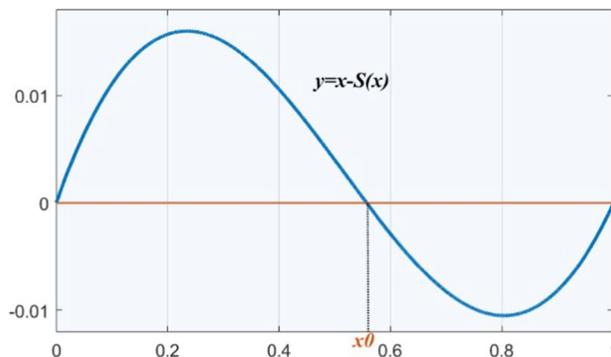
$$\sigma_i = \beta_i \cdot \max(\mathbf{I}), i = 1, 2, 3, \tag{4}$$

where  $\max(\mathbf{I})$  is the maximum of the image, and  $\beta_i$  is a positive constant (e.g., 0.8, 0.5 and 0.2).

After that, the estimation acquired from Eq. (3) is requisite to be adjusted via the intensity correction, for further improving the visual quality. According to the histogram of image, the intensity correction is,

$$\mathbf{R}'' = \begin{cases} 0, & \text{if } \mathbf{R}' \leq \mathbf{R}^{low} \\ \left( \frac{\mathbf{R}' - \mathbf{R}^{low}}{\mathbf{R}^{up} - \mathbf{R}^{low}} \right)^\gamma, & \text{if } \mathbf{R}^{low} \leq \mathbf{R}' \leq \mathbf{R}^{up} \\ 1, & \text{if } \mathbf{R}' \geq \mathbf{R}^{up} \end{cases}, \tag{5}$$

where  $\mathbf{R}^{up}$  and  $\mathbf{R}^{low}$  denote high and low shearing points of the image, and  $\gamma$  is a constant. We choose upper and lower confidence limits of confidence intervals as the high and low shearing points, in which the confidence level is set to 0.99 in the cumulative histogram in this section.



**Fig. 2** The graph of  $y = x - S(x), x \in [0, 1]$

### Backbone network

U-net and its variants (e.g., U-net ++ and U-net 3+), deeply-supervised encoder-decoder architecture, are widely applied for medical image segmentation with high efficiencies [14], because they can effectively capture fine-grained details of foreground objects. Moreover, ResNet is a powerful CNN, which helps the encoder in U-net become more effective [21]. Compared with U-net, U-net++ holds more skip connections that help the decoder acquire more information from multi levels of encoder [14]. Accordingly, ResUNet++ is developed to replace U-net with U-net++ and applies ResNet for the encoder [20], which achieves significant performance boost for a small number of images. Derived from those viewpoints, DFA-Net adopts ResUNet++ as its backbone semantic segmentation network. Even although CIET improves the contrast between the foreground objects (e.g., coronary vessels) and intricate background, interference factors (e.g., noise and other tissues with similar intensities) are also elevated to some extent, which has an effect on the segmentation. If original and improved images are individually input into the semantic network, dual multi-scale features are acquired accordingly that embed complementary information from different contrasts, in which the parameter sharing tactic is applied during training.

Since ResNet comprises five layers, termed as layer  $[j], j = 0, \dots, 4$ , and layer  $[k], k = 1, \dots, 4$ , consists of different number of double conv blocks, in which every block holds a  $3 \times 3$  conv, a batch normalization and a ReLU. As displayed in Fig. 1, the encoding process is described as follows. An input image  $x^i, i = 1, 2$ , with size of  $512 \times 512$ , passes through layer [0] that includes a  $7 \times 7$  conv with 64 filters and stride of 2, and then produces an output  $x_{0,0}^i$ . Next,  $x_{0,0}^i$  goes through layer [1] (three double conv blocks) of ResNet, and outputs  $x_{1,0}^i$ . After,  $x_{1,0}^i$  passes through layer [2] (four double conv blocks), and outputs  $x_{2,0}^i$ . Next,  $x_{2,0}^i$  goes through layer [3] (six double conv blocks), and produces  $x_{3,0}^i$ . Last,  $x_{3,0}^i$  passes through layer [4] (three double conv blocks), and yields  $x_{4,0}^i$ .

Details of the decoding process are described as follows. First, by upsampling  $x_{4,0}^i$  and concatenating with  $x_{3,0}^i$ , and then passing them through a triple conv block, we get  $x_{3,1}^i$ . Second, by upsampling  $x_{3,1}^i$  and concatenating with  $x_{2,0}^i$ , and then passing them through a triple conv block, we get  $x_{2,1}^i$ . Third, by concatenating  $x_{2,0}^i$  and  $x_{2,1}^i$ , and upsampling of  $x_{2,1}^i$ , and then pass them through a triple conv block, we get  $x_{2,2}^i$ . Fourth, by upsampling  $x_{2,2}^i$ , and concatenating with  $x_{1,0}^i$ , and then passing them through a triple conv block, we get  $x_{1,1}^i$ . Fifth, by concatenating  $x_{1,0}^i$  and  $x_{1,1}^i$  and upsampling of  $x_{1,1}^i$ , and then pass them through a triple conv block, we get  $x_{1,2}^i$ . Sixth, by concatenating  $x_{1,0}^i, x_{1,1}^i$  and  $x_{1,2}^i$ , and upsampling of  $x_{1,2}^i$ , and then passing them through a triple conv block, we get  $x_{1,3}^i$ . Seventh, by upsampling  $x_{1,3}^i$  and concatenating with  $x_{0,0}^i$ , and then passing them through a triple conv block, we get  $x_{0,1}^i$ . Eighth, by concatenating  $x_{0,0}^i$  and  $x_{0,1}^i$ , and upsampling of  $x_{0,1}^i$ , and then pass them through a triple conv block, we get  $x_{0,2}^i$ . Ninth, by concatenating  $x_{0,0}^i, x_{0,1}^i$  and  $x_{0,2}^i$ , and upsampling of  $x_{0,2}^i$ , and then passing them through a triple conv block, we get  $x_{0,3}^i$ . Last, by concatenating  $x_{0,0}^i, x_{0,1}^i, x_{0,2}^i$ , and  $x_{0,3}^i$ , and upsampling of  $x_{0,3}^i$ , and then passing them through a triple conv block, we get  $x_{0,4}^i$ .

### Feature aggregation

Since ResUNet++ separately accepts images with different contrasts as inputs, those corresponding multi-scale features are gained, that is,  $x_{0,i}^1$  and  $x_{0,i}^2$ ,  $i = 1, \dots, 4$ , which include different contrast information. After the feature aggregation that comprises the concatenation and addition under adaptive weighting parameters (see Fig. 1), dual multi-scale features are acquired, which produces a prediction map  $\mathbf{P}$  of pixel-wise probability for segmentation. The feature aggregation can be described as,

$$\mathbf{P} = \sum_{i=1}^4 \alpha_i \cdot \text{cat}(x_{0,i}^1, x_{0,i}^2), \quad (6)$$

where  $\text{cat}$  denotes concatenation, and  $\alpha_i$ ,  $i = 1, \dots, 4$ , denote adaptive weighting parameters.

During the back propagation process, the iteration of adaptive weighting parameters can be described as,

$$\begin{aligned} \alpha_k^{t+1} &= \alpha_k^t - \mu \frac{\check{s}^t}{\sqrt{\check{r}^t} + \varepsilon}, \check{s}^t = \frac{s^t}{1 - \rho_1^t}, \check{r}^t = \frac{r^t}{1 - \rho_2^t}, k = 1, \dots, 4, \\ s^t &= \rho_1^t s^{t-1} + (1 - \rho_1^t) g_k, r^t = \rho_2^t r^{t-1} + (1 - \rho_2^t) g_k, g_k = \nabla \hat{J}(\alpha_k^t), \end{aligned} \quad (7)$$

where  $g_k$  is the partial derivative of  $\hat{J}$  (loss function) with respect to  $\alpha_k$ ,  $s^t$  and  $r^t$  are the first and second moment estimates,  $\check{s}^t$  and  $\check{r}^t$  are the correction results of the first and second moment estimates,  $\mu$  is the learning rate,  $t$  is the iteration step,  $\rho_1$  and  $\rho_2$  are weight decay, and  $\varepsilon$  is a constant, avoiding the denominator of zero, respectively.

### Loss function

Since the vessel segmentation in X-ray angiography can be considered as a classification task, the concept of cross-entropy loss is adopted in this section because it is insensitive to the identity of assigned class in case of misclassifications [22]. However, misclassifying vessels (or foreground pixels) as ‘background’ (false negatives) is potentially at higher risk than misclassifying background pixels as ‘foreground’ (false positives). The reason maybe that the identification of vascular occlusion or stenosis is important in clinical diagnosis, while false negatives possibly result in missed opportunities of effective treatment at an early stage. Considering those different risks under different misclassification cases, a risk cross-entropy loss is introduced to address this issue, that is, to differentiate between false negatives and false positives by penalizing each error differently. For the sample  $x_i$ , label  $y_i$  and output of DFA-Net  $p_i$ ,  $i = 1, \dots, n$ , where  $n$  is the number of samples, the risk cross-entropy loss is defined as,

$$L_r = \frac{1}{n} \sum_{i=1}^N - [\gamma \cdot y_i \cdot \log p_i + (1 - y_i) \cdot \log(1 - p_i)], \quad (8)$$

where  $\gamma$  is a risk factor. If  $y_i > 0.5 + p_i$ ,  $\gamma > 1$ , else  $\gamma = 1$ .

Besides, Dice loss is commonly applied in semantic segmentation [15], which tells how well the model is performing when it comes to detecting boundaries regarding ground truth. Then, Dice loss is also adopted, which is defined as,

$$L_d = 1 - \frac{2 \sum_{i=1}^N z_i \hat{z}_i}{\sum_{i=1}^N z_i + \sum_{i=1}^N \hat{z}_i}, \quad (9)$$

where  $\hat{z}_i$  represents the prediction and  $z_i$  denotes ground truth.

From Fig. 1, it can be found that DSA images have high levels of noise, and with an imbalance between positive and negative samples. Because an adaptive moment estimation (Adam) strategy is computationally efficient, has little memory requirements, and is appropriate for non-stationary objectives and issues with noisy and/or sparse gradients [23], which is one of the most popular optimizer for accelerating the training of deep networks, DFA-Net is optimized with an Adam in this section, and the training process is implemented by minimizing the joint risk cross-entropy loss and Dice loss, which is defined as,

$$Loss = \alpha \times L_r + \beta \times L_d, \quad (10)$$

where  $\alpha$  and  $\beta$  are balance weightings. For each epoch in the training, the validation is used for evaluation regarding accuracy. The result with the highest accuracy is selected as the final model.

## Experimental results

### Metrics, baseline methods and datasets

Several widely-applied quantitative measures are adopted to evaluate the performance of different segmentation strategies, including two overlap-based metrics, e.g., intersection over union (IoU) and Dice coefficient (Dice), accuracy, specificity, and precision, which are defined as,

$$IoU = \frac{|X \cap Y|}{|X \cup Y|}, Dice = \frac{2 \cdot |X \cap Y|}{2 \cdot |X \cap Y| + |Y/X| + |X/Y|}, \quad (11)$$

$$Accuracy = \frac{|X \cap Y| + |\dot{X} \cap \dot{Y}|}{|X \cap Y| + |\dot{X} \cap \dot{Y}| + |Y/X| + |X/Y|}, \quad (12)$$

$$Specificity = \frac{|\dot{X} \cap \dot{Y}|}{|X/Y| + |\dot{X} \cap \dot{Y}|}, precision = \frac{|X \cap Y|}{|X \cap Y| + |X/Y|}, \quad (13)$$

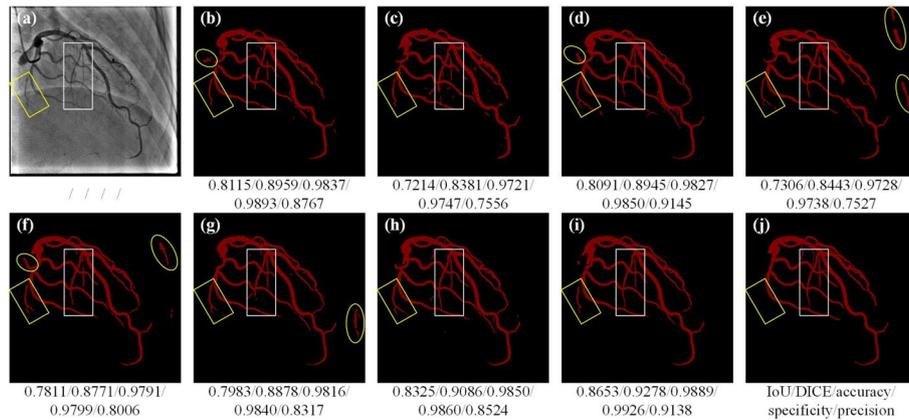
where  $X$  and  $Y$ , binary vectors (0 and 1 denote background and foreground areas), are the prediction and ground truth,  $|X \cap Y|$ ,  $|Y/X|$ ,  $|X/Y|$ , and  $|\dot{X} \cap \dot{Y}|$  are the number of true positives, false positives, false negatives, and true negatives, respectively.

In order to validate pros and cons of DFA-Net, seven deep learning-based segmentation strategies are selected as baselines for comparisons, e.g., U-net [13], Deeplabv3+ [24], U-net++ [14], SA-Unet [25], U-net 3+ [16], CMU-Net [26], and CA-Net [27], whose training models are re-trained for supplying optimal results for comparisons.

Besides that, after injected a contrast dye through the catheter, DSA provides temporal and/or spatial information that helps to visualize blood flow over time. However, we only apply DSA images under maximum filling of contrast dye to demonstrate the accuracy and superiority of DFA-Net. This work enrolled 78 participants, containing 48

**Table 1** Demographic characteristics of participants

	Number	Age (year)	Medical history					
			atherosclerosis	thoracalgia	dyspnea	palpitation	abnormal ECG	others
Male	48	59.7 ± 11.6	4	22	22	15	3	3
Female	30	62.3 ± 8.4	3	11	8	8	0	4
Total	78	60.7 ± 10.5	7	33	30	23	3	7



**Fig. 3** For a representative DSA image (a), segmented results obtained from U-net (b), Deeplabv3+ (c), U-net++ (d), SA-Unet (e), U-net 3+ (f), CMU-Net (g), CA-Net (h), and DFA-Net (i), respectively, and ground truth (j), in which white rectangles denote ROIs (e.g., tiny sub-branches), yellow ellipses indicate false detections, and yellow rectangles indicate capillary vessels that are not manually labelled in ground truth

male subjects (aged 22–77 years) and 30 female subjects (aged 51–77 years) from the Tongji Hospital, and the total number of images adopted was 292. Informed consent was acquired from each participant before DSA examination according to the procedure approved by the local ethics committee. Table 1 lists the demographic information for those participants, including age, gender, and medical history.

The vascular contour in each DSA image with size of  $512 \times 512$  is manually delineated by some experts, which is termed as ground truth (or label) during the training and testing. Under this condition, the dataset is randomly divided into 80% of training, 10% of validation, and 10% of testing. During training, the epoch is set to 400, batch size is set to 4, learning rate is set to  $10^{-4}$ , optimizer is chosen as Adam, the weight decay is initialized to  $10^{-4}$ , and loss function is chosen as the joint risk cross-entropy loss and Dice loss. Experiments are conducted using PyTorch 1.10.1 on a GPU (GeForce RTX 3090), and the operating system environment used is Linux.

**Qualitative comparisons**

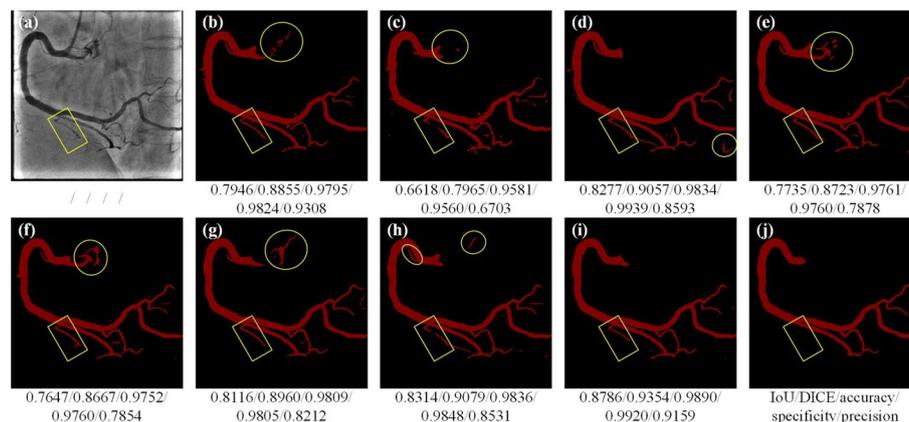
For a representative DSA image under maximum filling of contrast dye, Fig. 3 displays original image and segmented results obtained from baseline and proposed strategies, and ground truth (i.e., manual labels), and quantitative measures regarding the five metrics are presented below each image, in which yellow ellipses denote false detections, white rectangles indicate regions of interest (ROIs), e.g., tiny sub-branches, that are

important in the study of coronary blood circulation, and yellow rectangles suggest capillary vessels that are not manually labelled in ground truth (or considered as missed detections in ground truth). It can be seen that many false detections are yielded in those segmented results through U-net, U-net++, SA-Unet, U-net 3+, and CMU-Net, especially in Fig. 3e–g. Besides, the vessel integrity of ROIs (as displayed in white rectangles) is difficult to be preserved in Fig. 3b, c and f. The reason maybe that the contrast in those ROIs is inconspicuous, which causes some difficulties in the coronary vessel segmentation. If the contrast in ROIs is improved, the segmentation performance is promoted. This is a cornerstone of DFA-Net. By comparison, DFA-Net yields less false positives, and well protects the completeness of true positives, which implies that DFA-Net well matches ground truth.

Furthermore, yellow rectangles in Fig. 3a indicate capillary vessels, but they are not labelled in j. However, the baseline and proposed strategies identify those tiny sub-branches because of strong learning capabilities of deep models. This indicates that a good model could have a potential to supplement true detections to manual labels. If segmentation results acquired from an appropriate learning method are provided to experts for manual interactions (or correction) in prior, the annotation efficiency and accuracy will be significantly advanced. In this case, DFA-Net can competent for such task.

As for the quantitative measures, DFA-Net achieves the highest IoU, Dice, accuracy, and specificity scores, and the second-rank precision one. Compared with Deeplabv3+, the lift ratios (%) of DFA-Net are 19.95, 10.70, 1.73, 1.84, and 20.94 in terms of IoU, Dice, accuracy, specificity, and precision measures, respectively, and they become 8.39, 4.51, 0.74, 0.87, and 9.87, if compared with CMU-Net. Although U-net++ gains the best precision values, false positives and false negatives are noticeable. Both qualitative and quantitative comparisons from Fig. 3 demonstrate that DFA-Net achieves superior fidelity and structure similarity for those segmentation results.

For another representative DSA image under maximum filling of contrast dye, Fig. 4a denotes the original image, and Fig. 4b–i display segmented results acquired from U-net,



**Fig. 4** For an original DSA image (a), (b)–(i) denotes segmented results obtained from U-net, Deeplabv3+, U-net++, SA-Unet, U-net 3+, CMU-Net, CA-Net, and DFA-Net, respectively, and (j) is ground truth, where yellow ellipses denote false detections, and yellow rectangles indicate capillary vessels that are not manually labelled in ground truth

Deeplabv3+, U-net++, SA-Unet, U-net 3+, CMU-Net, CA-Net, and DFA-Net, respectively, and Fig. 4(j) is ground truth, where yellow ellipses denote false detections and yellow rectangles denote capillary vessels that are not manually labelled in ground truth. Similar to Fig. 3, some false detections are found in the results through baseline methods, with different degrees of distortion. In contrast, DFA-Net gains less false detections, which is helpful for early detection of cardiovascular diseases. Besides, DFA-Net achieves the highest IoU, Dice, accuracy, and precision scores, and the second-best specificity one. When compared with U-net 3+, the lift ratios (%) of DFA-Net are 14.89, 7.93, 1.42, 1.64, and 16.62 regarding those five metrics, and they become 13.59, 7.23, 1.32, 1.64, and 16.26, if compared with SA-Unet. The comparisons from Figs. 3 and 4 validate the superiority of DFA-Net over baseline tactics.

### Quantitative comparisons

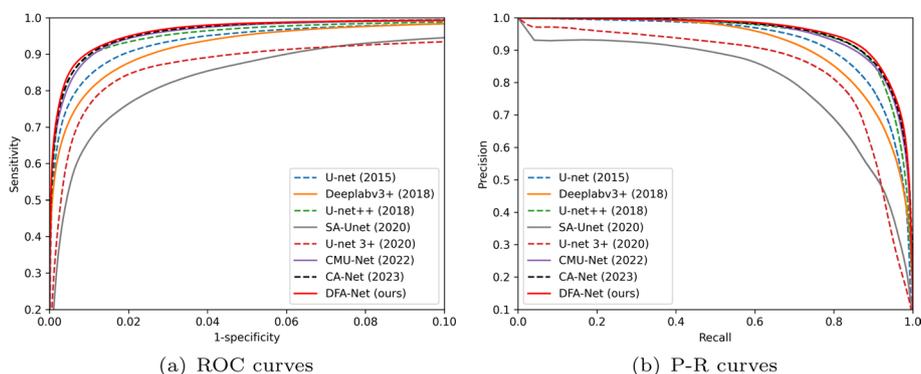
Since IoU, Dice, accuracy, specificity, and precision are widely-applied to measure segmentation performance, higher values suggest better performance. Table 2 lists quantitative comparisons for the testing dataset acquired from U-net, Deeplabv3+, U-net++, SA-Unet, U-net 3+, CMU-Net, CA-Net, and DFA-Net regarding the five metrics. It can be found that DFA-Net achieves the best results in terms of each metric, CA-Net acquires the second-rank IoU, Dice and accuracy scores, and U-net++ acquires the second-rank specificity and precision values. Compared with second-rank results, the lift ratios of DFA-Net are 1.65, 0.92, 0.15, 0.25, and 3.35 (%) regarding IoU, Dice, accuracy, specificity, and precision, respectively. If compared with Deeplabv3+, the respective lift ratios of DFA-Net are 19.02, 10.56, 1.47, 1.62, and 21.96, and they become as 3.42, 1.90, 0.31, 0.52, and 6.71 (%), when compared with CMU-Net. This indicates that DFA-Net is superior to the seven baseline approaches regarding each measure, which is consistent with the conclusions derived from Figs. 3 and 4. In this case, DFA-Net holds high reliability and robustness for the segmentation of DSA images, which is instructive to highlight ROIs, understand context, and represent structure details of images.

A receiver operating characteristic (ROC) curve represents the sensitivity and specificity pairs corresponding to a particular decision threshold, and the area under ROC curve (AUC) is a metric of how well a parameter identifies between two diagnostic groups. Besides that, an alternative to ROC curve is the precision-recall (P-R) curve that is considered as a supplement to ROC curve when evaluating and comparing tests because it

**Table 2** Quantitative comparisons acquired from different segmentation strategies

Method	IoU	Dice	Accuracy	Specificity	Precision
U-net (2015)	0.7377	0.8491	0.9798	0.9896	0.8538
Deeplabv3+ (2018)	0.6731	0.8046	0.9709	0.9766	0.7331
U-net++ (2018)	0.7852	0.8797	0.9836	<b>0.9899</b>	<b>0.8651</b>
SA-Unet (2020)	0.7201	0.8373	0.9767	0.9829	0.7894
U-net 3+ (2020)	0.7541	0.8598	0.9800	0.9848	0.8123
CMU-Net (2022)	0.7746	0.8730	0.9822	0.9873	0.8379
CA-Net (2023)	<b>0.7881</b>	<b>0.8815</b>	<b>0.9837</b>	0.9896	0.8623
DFA-Net (ours)	<i>0.8011</i>	<i>0.8896</i>	<i>0.9852</i>	<i>0.9924</i>	<i>0.8941</i>

The best two results are marked in italic and bold



**Fig. 5** ROC curves **a** and P-R curves **b** obtained from different segmentation strategies

**Table 3** AUC and AP measures acquired from different segmentation algorithms

	U-net	Deeplabv3+	U-net++	SA-Unet	U-net 3+	CMU-Net	CA-Net	DFA-Net
AUC	0.9891	0.9904	0.9926	0.9748	0.9737	0.9946	<b>0.9952</b>	<i>0.9953</i>
AP	0.9240	0.9126	0.9458	0.8054	0.8484	0.9495	<b>0.9527</b>	<i>0.9547</i>

The best two results are marked in italic and bold

may be a better choice for imbalanced data [28]. P-R curve represents the recall and precision pairs, and accepts the area under P-R curve (or named average precision, AP) as a performance metric. The ROC and P-R curves acquired from the baseline and proposed tactics are displayed in Fig. 5, and the corresponding AUC and AP values are listed in Table 3. It can be found that CA-Net acquires the second-best results, while DFA-Net achieves the highest AUC of 0.9953 and highest AP of 0.9547, which is consistent to the conclusions derived from Figs. 3 and 4, and Table 2. Accordingly, ROC and P-R curves validate that DFA-Net is able to distinguish coronary vessels from complex background, and construct an effective biomarker that decodes the characteristics of DSA images.

**Ablation study**

The proposed strategy integrates CIET (a contrast improvement module) into ResU-net++ (a semantic segmentation module), in which original and improved images are fed into the model separately, and the loss is chosen as joint risk cross-entropy loss and Dice loss. In order to demonstrate the contributions of each component, ablation experiments are implemented in terms of IoU, Dice, accuracy, specificity, and precision metrics. Eight configurations that possibly impact the segmentation performance are considered, e.g., w/ or w/o original / contrast-improved image, risk cross-entropy and Dice loss, as listed in Table 4. Besides, we analyze the compatibility with other loss functions, e.g., cross-entropy loss and focal loss.

Table 4 provides the ablation experimental results by disabling improved image, original image, risk cross-entropy loss, and Dice loss components to the baseline architecture (i.e., DFA-Net) on the testing data regarding IoU, Dice, accuracy, specificity, and precision measures. It can be found that the baseline acquires the best quantitative

**Table 4** Quantitative analyses according to different configurations

Configuration				Metric				
Original	Improved	Risk loss	Dice loss	IoU	Dice	accuracy	specificity	precision
✓		✓	✓	<b>0.7883</b>	<b>0.8816</b>	<b>0.9842</b>	0.9918	0.8849
	✓	✓	✓	0.7821	0.8778	0.9837	0.9916	0.8824
✓	✓		✓	0.7715	0.8710	0.9830	<b>0.9923</b>	<b>0.8902</b>
✓	✓		✓ <sub>f</sub>	0.7502	0.8573	0.9790	0.9821	0.7903
✓	✓	✓		0.7723	0.8715	0.9824	0.9893	0.8565
✓	✓	✓ <sub>f</sub>	✓	0.7843	0.8791	0.9838	0.9917	0.8834
✓	✓	✓	✓ <sub>f</sub>	0.7668	0.8680	0.9812	0.9855	0.8206
✓	✓	✓	✓	<i>0.8011</i>	<i>0.8896</i>	<i>0.9852</i>	<i>0.9924</i>	<i>0.8941</i>

The best two results are marked in italic and bold ✓<sub>f</sub> denotes that cross-entropy loss substitutes for risk cross-entropy loss, and ✓ denotes that focal loss substitutes for Dice loss in the proposed model

results among the eight configurations, which indicates the incorporation of original and improved images, risk cross-entropy and Dice loss improves the segmentation performance of DSA images. If removing one or more components, the performance is degraded. For example, if only original image is input into DFA-Net, the performance is slightly superior to that of the organization that only accepts contrast-improved image as an input. Nonetheless, the concatenation of extracted features derived from original and contrast-improved images is able to evidently boost the metrics. The reason maybe that images with different contrasts embed complementary information, which is helpful to strengthen learning capabilities of networks, and thus effectively segment coronary vessels.

From Table 4, it can be also found that the risk cross-entropy loss makes more contributions to DSA segmentation regarding IoU and Dice measures, but a trifle inferiority in terms of other terms, in comparison with the Dice loss. Nevertheless, the joint of risk cross-entropy and Dice loss holds superior performance regarding the five terms. Besides that, other loss functions, e.g., cross-entropy loss and focal loss, have their own pros and cons. These suggest that appropriate loss functions guarantee good segmentation performance of DSA images.

In general, different configurations possess advantages and disadvantages, which demonstrates the importance and necessity of each component incorporated in DFA-Net. Accordingly, those ablation experimental results demonstrate that DFA-Net is dependable and impactful to segment coronary vessels in DSA images.

**Discussion**

In recent years, deep learning is fascinated in fields of angiographic segmentation and achieves prominent progresses, while most of tactics bring high false and missed detections due to indistinct contrast between vessels and background, especially on tiny sub-branches. Image improvement strategies better such contrast, while boosting extraneous information, e.g., catheter or other tissues with similar intensities. An incorporation of complementary information derived from diverse contrasts is able to strengthen learning abilities of networks. Accordingly, inspired by advantages of contrast improvement

and encoding-decoding architecture, DFA-Net is proposed, incorporating a contrast improvement using exponent transformation (i.e., CIET) into a semantic segmentation network (i.e., ResUnet++). Meanwhile, the joint risk cross-entropy loss and Dice loss are imposed on DSA segmentation. Compared with state-of-the-art tactics, DFA-Net achieves superior performance with regard to IoU, Dice, accuracy, specificity, and precision measures. This indicates that DFA-Net holds high fidelity and robustness to ground truth, providing an active way in the interpretation of DSA images.

Even although DFA-Net yields promising segmentation results under different cases, there are some limitations to this work. Firstly, the number of DSA images and labels is insufficient. Nevertheless, large amounts of training data are required to train reliable and effective deep learning models. Therefore, the further work is to collect massive data with manual labels. Secondly, although the encoder-decoder structure is considered as the backbone, other advanced networks (e.g., unsupervised and semi-supervised models) are compatible in the future. Thirdly, the compatibility with other loss functions (e.g., focal loss and cross-entropy loss) is discussed in the methodology (see Table 4), but other advanced loss functions, e.g., direction connection loss and Tversky loss, will be exploited in the future. Lastly, the temporal and spatial domain information is insufficiently explored, which leads to a future work.

## Conclusion

In order to overcome difficulties of angiographic segmentation caused by indistinct contrast between vessels and background, especially on tiny sub-branches, a dual multi-scale feature aggregation network is proposed in this work (i.e., DFA-Net), which respectively accepts original images and contrast-improved images as inputs because different contrasts embed complementary information that is advantageous to strengthen learning capabilities of networks, and thus obtain an available segmentation of coronary vessels. Meanwhile, a risk cross-entropy loss is imposed on the segmentation, for effectively lessening false negatives, which is incorporated with Dice loss for jointly optimizing the proposed model during training. Experimental results validate that DFA-Net not only works more effectively for diverse DSA images, but also has significant superiorities over state-of-the-art tactics, e.g., the highest IoU, Dice, accuracy, specificity, and precision measures. This indicates that DFA-Net achieves high effectiveness and robustness to ground truth. Consequently, DFA-Net is promising in the coronary vessel segmentation of X-ray angiography, and facilitates the study in the explanation and understanding of ROIs in DSA images. This leads to a linchpin for our further work.

## Acknowledgements

Not applicable.

## Author contributions

HD and XL designed and implemented the proposed model; XL, YQL and TF supervised the methodology and tests. XDM provided the datasets. All authors read and approved the final manuscript.

## Funding

This work was supported in part by National Natural Science Foundation of China (62071456, 82102028), "The 14th Five Year Plan" Hubei Provincial advantaged characteristic disciplines (groups) project of Wuhan University of Science and Technology (2023D0502), and Hubei Province Key Laboratory of Systems Science in Metallurgical Process (Y202305).

## Data availability

Not applicable.

## Declarations

### Ethics approval and consent to participate

The authors accept the Journal of Big Data's ethics approval and agree to share their work for scientific advancement.

### Competing interests

The authors disclose that they do not have any Conflict of interest.

Received: 11 December 2023 Accepted: 14 March 2024

Published online: 26 April 2024

## References

- Cardiovascular diseases (CVDs). <https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-cvds>.
- Tao X, Dang H, Zhou X, Xu X, Xiong D. A lightweight network for accurate coronary artery segmentation using x-ray angiograms. *Front Pub Health*. 2022;10:892418. <https://doi.org/10.3389/fpubh.2022.892418>.
- Myerburg RJ, Junttila MJ. Sudden cardiac death caused by coronary heart disease. *Circulation*. 2012;125(8):1043–52. <https://doi.org/10.1161/CIRCULATIONAHA.111.023846>.
- Gao Z, Wang L, Sorousmehr SMR, Wood A, Gryak J, Nallamothu B, Najarian K. Vessel segmentation for x-ray coronary angiography using ensemble methods with deep learning and filter-based features. *BMC Med Imaging*. 2022;22:1–17. <https://doi.org/10.1186/s12880-022-00734-4>.
- Wang W, Xia Q, Yan Z, Hu Z, Chen Y, Zheng W, Wang X, Nie S, Metaxas D, Zhang S. Avdnet: Joint coronary artery and vein segmentation with topological consistency. *Med Image Anal*. 2024;91:102999. <https://doi.org/10.1016/j.media.2023.102999>.
- Pu Y, Zhang Q, Qian C, Zeng Q, Li N, Zhang L, Zhou S, Zhao G. Semi-supervised segmentation of coronary dsa using mixed networks and multi-strategies. *Comput Biol Med*. 2023;156:106493. <https://doi.org/10.1016/j.combiomed.2022.106493>.
- Zhou H, Xiao J, Li D, Fan Z, Ruan D. Intracranial vessel wall segmentation with deep learning using a novel tiered loss function to incorporate class inclusion. In: 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI); 2022. p. 1–4.
- Wan T, Chen J, Zhang Z, Li D, Qin Z. Automatic vessel segmentation in x-ray angiogram using spatio-temporal fully-convolutional neural network. *Biomed Signal Process Control*. 2021;68:102646. <https://doi.org/10.1016/j.bspc.2021.102646>.
- Gharleghi R, Chen N, Sowmya A, Beier S. Towards automated coronary artery segmentation: a systematic review. *Comput Methods Programs Biomed*. 2022;225:107015. <https://doi.org/10.1016/j.cmpb.2022.107015>.
- Shen N, Xu T, Huang S, Mu F, Li J. Expert-guided knowledge distillation for semi-supervised vessel segmentation. *IEEE J Biomed Health Inf*. 2023;27(11):5542–53. <https://doi.org/10.1109/JBHI.2023.3312338>.
- Gao Z, Zong Q, Wang Y, Yan Y, Wang Y, Zhu N, Zhang J, Wang Y, Zhao L. Laplacian salience-gated feature pyramid network for accurate liver vessel segmentation. *IEEE Trans Med Imaging*. 2023;42(10):3059–68. <https://doi.org/10.1109/TMI.2023.3273528>.
- Zhang H, Gao Z, Zhang D, Hau WK, Zhang H. Progressive perception learning for main coronary segmentation in x-ray angiography. *IEEE Trans Med Imaging*. 2023;42(3):864–79. <https://doi.org/10.1109/TMI.2022.3219126>.
- Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: *medical image computing and computer-assisted intervention – MICCAI 2015*; 2015. p. 234–241.
- Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. Unet++: A nested u-net architecture for medical image segmentation. *Deep Learn Med Image Anal Multimodal Learn Clin Decis Support*. 2018;2018(11045):3–11.
- Oktay O, Schlemper J, Folgoc L, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla N, Kainz B, Glocker B, Rueckert D. Attention u-net: learning where to look for the pancreas. In: *medical imaging with deep learning, (MIDL)*; 2018.
- Huang H, Lin, L, Tong R, Hu H, Zhang Q, Iwamoto Y, Han X, Chen Y-W, Wu J. Unet 3+: A full-scale connected unet for medical image segmentation. In: *ICASSP 2020 - 2020 IEEE international conference on acoustics, speech and signal processing (ICASSP)*; 2020. p. 1055–1059.
- Dong C, Xu S, Dai D, Zhang Y, Zhang C, Li Z. A novel multi-attention, multi-scale 3d deep network for coronary artery segmentation. *Med Image Anal*. 2023;85:102745. <https://doi.org/10.1016/j.media.2023.102745>.
- Shi T, Ding X, Zhou W, Pan F, Yan Z, Bai X, Yang X. Affinity feature strengthening for accurate, complete and robust vessel segmentation. *IEEE J Biomed Health Inf*. 2023;27(8):4006–17. <https://doi.org/10.1109/JBHI.2023.3274789>.
- Jobson DJ, Rahman Z, Woodell GA. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Trans Image Process*. 1997;6(7):965–76. <https://doi.org/10.1109/83.597272>.
- Zhao T, Pan S, He X. Resunet++ for sparse samples-based depth prediction. In: 2021 IEEE 15th international conference on electronic measurement & instruments (ICEMI); 2021. p. 242–246.
- He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: 2016 IEEE conference on computer vision and pattern recognition (CVPR); 2016. p. 770–778.
- Xie Y, Xia Y, Zhang J, Song Y, Feng D, Fulham M, Cai W. Knowledge-based collaborative deep learning for benign-malignant lung nodule classification on chest ct. *IEEE Trans Med Imaging*. 2019;38(4):991–1004. <https://doi.org/10.1109/TMI.2018.2876510>.
- Kingma D, Ba J. Adam: a method for stochastic optimization. In: *international conference on learning representations*; 2014.
- Chen L-C, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: *computer vision – ECCV 2018*; 2018. p. 833–851.

25. Guo C, Szemenyei M, Yi Y, Wang W, Chen B, Fan C. Sa-unet: Spatial attention u-net for retinal vessel segmentation. In: 2020 25th international conference on pattern recognition (ICPR); 2021. p. 1236–1242.
26. Tang F, Wang L, Ning C, Xian M, Ding J. Cmu-net: A strong convmixer-based medical ultrasound image segmentation network. In: 2023 IEEE 20th international symposium on biomedical imaging (ISBI); 2023. p. 1–5.
27. Xie X, Zhang W, Pan X, Xie L, Shao F, Zhao W, An J. Canet: context aware network with dual-stream pyramid for medical image segmentation. *Biomed Signal Process Control*. 2023;81:104437. <https://doi.org/10.1016/j.bspc.2022.104437>.
28. Wasikowski M, Chen X-W. Combating the small sample class imbalance problem using feature selection. *IEEE Trans Knowl Data Eng*. 2010;22(10):1388–400. <https://doi.org/10.1109/TKDE.2009.187>.

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.